The OA Interviews: ScienceOpen's Alexander Grossmann

RICHARD POYNDER, 16th November 2015

To go straight to the interview, click here

In his time, the founder and president of ScienceOpen, Alexander Grossmann, has sat on both sides of the scholarly publishing table. He started out as a researcher and lecturer, working variously at the Jülich Research Centre, the Max Planck Institute in Munich and the University of Tübingen.

Then in 2001 he reinvented himself as a publisher, working first at Wiley-Blackwell, and subsequently as managing director at Springer-Verlag GmbH in Vienna, and a vice president at De Gruyter.

An important moment for Grossmann came in 2008, when Springer acquired the open-access publisher BioMed Central from serial entrepreneur Vitek Tracz. Listening to a presentation on the purchase given at a management meeting by the company's CEO Derk Haank, Grossmann immediately saw the logic of the move, and the imperatives of open access.

However, it was soon apparent to him that the publishing industry at large is not in a hurry to reinvent itself for an OA world, and certainly not if it means having to take hard decisions that could threaten the high profit levels that it has become accustomed to earning from journal publishing.

Speaking to me two years ago Grossmann put it this way: "[T]here is no publishing house which is either able or willing to consider the rigorous change in their business models which would be required to actively pursue an open access publishing concept."

And this remains his view today.

In 2013, therefore, Grossmann partnered with Boston-based entrepreneur and software developer Tibor Tscheke to found a for-profit OA venture called ScienceOpen. At the same time he took a post as professor of publishing management at the Leipzig University of Applied Sciences.

It is important to note that ScienceOpen is not just another OA publisher. While it does offer scholarly publishing services, its vision and *modus operandi* is not the same as most OA publishers. Not only does it make all the papers it publishes open access, but in most cases the reviewing takes place after publication rather than before, with the reviews made openly available too. In addition, the reviews have the names of the reviewers attached to them. ScienceOpen practises what Grossmann calls "non-anonymous public Post-Publication Peer Review", or PPPR.

But what is most distinctive about ScienceOpen is that it is also harvesting papers from other publishers and web sites. To date it has imported 950,000 open access articles from the PubMed Central Open Access Subset and 830,000 articles from the physics pre-print server arXiv. It has also harvested the bibliographic details of 8 million papers published in subscription journals. In doing all this its objective is not just to offer a new search tool for academia, but to encourage users to comment on and review the papers it has aggregated.

As a further innovation, ScienceOpen has begun to create "collections" of articles on specific topics, with papers aggregated from different sources and brought together in a thematic way. Currently there are 15 such collections. ScienceOpen is also deploying new recognition and reputation systems

like [Altmetric](#) on its site, with alternative metrics data available for all of ScienceOpen's nearly 10 million article records.

In light of its approach, ScienceOpen describes itself not as a publisher, but a "research and publishing network". The aim, it says, is to allow researchers to share scientific information with one another, both formally by publishing articles, and informally by reviewing their colleagues' work, providing endorsements and comments, and updating their own papers.

Or as Grossmann puts it, ScienceOpen consists of "three core services: publishing, aggregation and engagement, all of which are "rooted in the principles of open access and transparency."

Some background

Inherent to the creation of the ScienceOpen venture is a belief that if scholarly publishing is to be effectively disrupted it is necessary to do more than simply launch a few new OA journals and replicate the traditional customs and practices of legacy publishers.

As Grossmann explains, "[E]very year about 2 million new articles are published in STM alone (out of 4 million submitted manuscripts). So even at its most successful, a new for- or non-profit venture can expect to have published only a few thousand to 10,000 papers after, let's say, five years. This is neither a significant market share, nor does it contribute to the disruptive change in the publishing industry that we can expect to see take place, and which has just started."

In other words, the revolution in scholarly publishing that OA advocates have long been calling for will require something more radical than we have seen to date.

Far from disrupting the market, in fact, open access and the digital revolution have only intensified a long-standing and unhealthy consolidation in the industry, which is now dominated by what has been termed an "[Oligopoly of Academic Publishers](#)". Moreover, membership of this powerful group consists exclusively of legacy publishers, including [Reed Elsevier](#), [Wiley-Blackwell](#), [Springer Nature](#), and [Taylor & Francis](#). Consequently, the top five most prolific publishers account for more than 50% of all the papers that were published in 2013.

What has been learned is that the research community's continuing addiction to the [Impact Factor](#) has enabled legacy publishers to appropriate the revolution that OA was meant to usher in and leverage it for their own purposes. This has not only seen their market power grow, but their revenues increase as well.

By introducing [hybrid open access](#), for instance, they have been able to continue charging subscription fees for their journals, while also levying article-processing charges (APCs) on a per-paper basis for those authors willing/required to make their work OA. This has given hybrid journals an additional revenue stream on top of their exorbitant subscriptions, a phenomenon generally referred to as "[double dipping](#)". Moreover, journals with a high IF/brand value are able to charge higher APCs than gold OA publishers. This gives legacy journals a further advantage as they have had time to acquire an Impact Factor.

As a result, many have come to conclude that scholarly publishing will never be truly disrupted unless the much maligned (and widely [discredited](#)) Impact Factor is jettisoned, and the 350-year-old scholarly journal significantly subverted, or even abandoned altogether.

That would certainly seem to be the logic of ScienceOpen's approach. By aggregating papers from many different publishers, for instance, it hopes to shift the emphasis from the journal to the article, and to replace the Impact Factor with article-level assessment and quality metrics.

Thus, as well as seeking to make the reviewing and commenting of research papers more open and transparent, ScienceOpen wants to see it flourish independently of the originating publishers, by centralising the process on its own platform. The hope is that this will see the research community begin to shift its focus from journals and publishers, to the article and the author.

The aim, says Grossmann, is to "start to de-couple where a researcher publishes from what they publish, and how their research is judged". This in turn, he says, can "loosen the vice-like grip of the Impact Factor."

Some comparisons

To put it another way, ScienceOpen wants to layer an open reviewing process over the top of the research corpus, and in the process weaken the brand value of journals and moot journal-level quality measures like the Impact Factor. If such a strategy were successful it could theoretically turn publishers into simple content providers more or less indistinguishable from one another, rather than must-have brands able to charge what they wish.

It is worth comparing what ScienceOpen is doing with the way in which overlay journals select, review, and aggregate papers deposited in arXiv — as Cambridge mathematician Timothy Gowers is currently planning to do with the *Discrete Analysis* journal.

However, where overlay journals tend to harvest papers before they have been published, ScienceOpen is aggregating them after they have been published. Moreover, it is seeking to aggregate not just preprints in mathematics, physics, astronomy, and computer science, but potentially the entire research corpus.

A further important distinction is that while an overlay journal may simply consist of a list of links to arXiv preprints, it will likely be entirely conventional in all other ways.

Another comparison we could make is with the social networking site Academia.edu, which by encouraging users to share papers with one another has become an unofficial research distribution channel for papers published elsewhere. Recently it also announced plans to layer a peer review service called PaperRank on top of the service. As such, it will provide the same reviewing capabilities as ScienceOpen. However, an important difference is that it has 27 million users (compared to ScienceOpen's 85,000) and 7.5 million papers on its site. Moreover, it claims to have signed up 2,000 editors to do the reviewing.

We should also note that while ScienceOpen has aggregated millions of article records, the vast majority of these consist of bibliographic details alone. The papers themselves are still sitting behind paywalls, and so invisible to anyone without subscription access to them. In addition, as publishers generally disallow green OA papers to be posted to centralised for-profit services, ScienceOpen is not harvesting self-archived papers. It seems highly unlikely that anyone would comment on or review a paper where the full-text is absent, not to say somewhat difficult!

Moreover, in some cases the details of the papers indexed by ScienceOpen are somewhat inadequate with, for instance, the article title and/or names of the authors entirely absent (e.g. here and here).

So how is ScienceOpen performing a year after launch? It has to be acknowledged that it does not appear to be a hive of activity so far as publishing and reviewing is concerned. To date it has published only 55 papers and 80 posters, and the site has attracted just 100 reviews.

In light of this low demand for its publishing services, ScienceOpen is now focusing more on its aggregation and engagement services. So, for instance, it is enhancing its "collection" features by adding new topics and editors, and providing new metrics and presentation options.

And in September it announced that it is creating a new open and freely available citation index. When complete this will list for each record in the database the number of citing articles (but only for those available in ScienceOpen), related articles, the most cited references and the most referenced authors.

But as ScienceOpen's emphasis shifts, the need for a more drastic intervention in scholarly publishing might appear to be that much greater — certainly if the market is to be truly disrupted. What is needed, perhaps, is the kind of intervention envisaged by F1000Research founder Vitek Tracz.

Speaking to me in September, Tracz said: "We believe that the main problem is the very existence of journals, and the methods their editors use to accept or reject what to publish. Their reason for making these choices is significantly affected by their battle to maintain and raise their Impact Factor."

Tracz's proposed solution is for funders and research institutions to bypass the traditional journal/publisher all together and start publishing their own research. As he put it, "Our hope is to convince funders and institutions to operate their own 'publishing platforms' and encourage (and in time perhaps mandate) the research they fund to be published on these platforms."

And Tracz is offering to license them the F1000Research publishing platform to enable them to do this.

Who, when and how?

But how likely is it that new services like ScienceOpen, Academia.edu and F1000Research can succeed in their aims? Apart from anything else, can they remain financially viable long enough to usher in the revolution they believe is needed? Having published just 55 papers (many of which will have had the $400 APC waived) and 80 posters (which attract no publishing fee), it is hard to how ScienceOpen's publishing operations can support its business.

What about its other activities? Can they generate sufficient revenues? When I raised the question of business models with Grossmann he said merely that the new citation index would be funded by "modest indexing fees and marketing opportunities for publishers. It legitimises content sharing and increases traffic to their content."

Presumably the plan is to charge open access publishers to co-host and promote their content — the kind of deal I assume has been done with Thieme, several of whose journals have recently been loaded on to ScienceOpen.

What about the long-term viability of Academia.edu? While it has attracted over $17 million in venture capital and gained a very large number of users, it is not clear to me what its business model is. After all, however much VC money a company may attract it will eventually run out.

The founder, Richard Price, has suggested that the company will make money by selling data about the content it hosts (see here and here for instance). However, when I emailed Price to ask about the company's finances he was tight-lipped, saying merely, "as a private company we don't comment on our finances".

To complicate the picture, Academia.edu's *modus operandi* is somewhat controversial, and subject to take-down notices (and presumably potential liability issues) from legacy publishers.

What about F1000Research? A web search on the company turned up the abbreviated accounts for year-end 2014. These indicate that net liabilities at the time stood at -£2.3 million (compared to -£940,000 the previous year).

A great many people in the research community are now impatient to see scholarly publishing disrupted and reinvented, and they assume that open access will be at the heart of that revolution. But who will engineer the revolution, and when and how it will take place has yet to be established.

For his part, Tracz has no illusions about the difficulty of the task ahead. Fixing the system, he told me, is "both the hardest, and the most important issue to solve."

For Alexander Grossmann's insights into these matters, and further information about ScienceOpen, please read the Q&A below.

The interview begins …

Alexander Grossmann

RP: ScienceOpen describes itself as "a freely accessible research network to share and evaluate scientific information". What does that mean in practice, and how does ScienceOpen differentiate itself from what traditional publishers do, and from what other OA publishers do?

AG: ScienceOpen is different because we try to open up all research, regardless of publisher, to post-publication evaluation — both through peer review and through community curation in article collections.

So a key difference between ScienceOpen and other open access publishing ventures is that ScienceOpen provides access not only to content that has been published on our site but also to nearly 2 million open access articles and to another 8 million bibliographic records that we have harvested.

All these articles can be read, commented upon or reviewed on ScienceOpen. And this is not just a fancy add-on feature: I strongly believe that the concept of a publishing platform that offers access only to content that it has published itself — as has been the norm for decades within the academic publishing industry — can no longer work.

We also strongly believe that it is important to create new mechanisms to allow researchers to build recognition and reputation after publication. By doing so we can start to de-couple where a researcher publishes from what they publish, and how their research is judged. And in this way we can hope to loosen the vice-like grip of the Impact Factor.

What we have in common with traditional publishers is that we also produce article collections in topical groupings. However, the difference is that rather than professional journal editors curating the content (pre and/or post publication), we empower the community (post-publication) to act as "Collection Editors". And we offer a nifty widget to enable them to do this. This is important because they are the subject experts, and so should be empowered to play this role in the publication process.

Open access publishing network

RP: What has ScienceOpen achieved over the past year, and how have your views changed over the past 12 months with regard to what it should be doing, and what services it should be offering?

AG: When I started to concretely think about the future of publishing and academic discourse four or five years ago, I had in mind an open access publishing *network*, and one that was different from the existing ventures in academic publishing.

My partner Tibor Tscheke and I, therefore, decided to start from the beginning with an open access publication network that would enable researchers to search, spot, discover and discuss relevant research, wherever it has been published.

When ScienceOpen launched in beta in November 2013 it already combined these elements. A year ago we also began to offer publishing services for authors who are interested in making their own research public immediately, and who understand that they will benefit from an open discussion of their work on ScienceOpen, enabled by non-anonymous public Post-Publication Peer Review (PPPR).

To this end we launched the journal *ScienceOpen Research* (ISSN 2199-1006), and introduced an exclusively post-publication review process. We have published over 50 articles so far, and have seen a lively discussion around each one. We have also launched the journal *ScienceOpen Posters* (ISSN 2199-8442) which to date has published 80 posters.

From the very beginning, ScienceOpen was open to publishing all types of research: articles, reviews, letters, comments, opinions, case studies, clinical reports, negative results, and posters.

In December 2014, the Max Planck organisation began a collaboration with ScienceOpen. As a result, all researchers affiliated with one of the 83 Max Planck institutions in Europe can publish their research on our platform at no cost.

Nonetheless, we view these publishing services as only a necessary part of our product portfolio. It is available to authors who demand it, but it is not the only element of our business model.

Over the past year it has become clear to us that the strength of ScienceOpen lies in our aggregation tools. Going forward, therefore, our focus will be on developing new ways to expose the relationships and context between articles and to improve the site's discoverability and capabilities for users. Much of that focus will be on expanding our Collection features with new topics and editors, new metrics and presentation options.

It has also become obvious to us that this can only be viewed as a single step towards creating the effective system for scholarly discourse that will be needed in the future. Researchers are going to rely less on the specific brands of publishers and journals, and focus instead on keywords and topics that are relevant to their research.

In the past 12 months, therefore, our database has grown from 500,000 open access articles to over 10 million articles and records. And in order to support researchers in discovering and filtering relevant articles from all publishers, we have taken the first step towards a publically available citation network. This is designed to enable researchers to sort results by relative citation numbers.

We will also shortly announce the strengthening of an important partner relationship leading to another major improvement in our service.

For more on ScienceOpen's vision I would refer you to my recent publication on the transition scholarly publishing needs to go through, which is available here.

Peer review

RP: You said that ScienceOpen practises what you call Post-Publication Peer Review. Perhaps you could say more about this, and how the process differs from other OA publishers like F1000Research and PeerJ?

AG: Basically, the ScienceOpen peer review concept is to publish first and then evaluate the work post-publication in a network context.

We believe that this, coupled with our integration with ORCID, is key to maintaining a higher standard of scientific discourse. Anyone with the required qualifications (5 peer reviewed

publications in ORCID) may review an article. In addition, anyone can invite a reviewer. Importantly, the identity of the reviewers is openly available to all users of ScienceOpen, who can access the full profile and ORCID record of the reviewer to make their own judgement about the relevance of the statements in the reviewers' report.

All of the newer open access publishing platforms and megajournals — e.g. F1000Research, PeerJ, and The Winnower — have introduced a more open concept of peer review, but there are some differences.

Thus far, The Winnower has the closest reviewing model to ours, which is also based on public PPPR. Other publishing platforms, such as F1000Research, have an editor-mediated process where they publish submissions first and then *invite* specific reviewers to submit a report on a paper, which is made public (Open PPPR).

In comparison, PeerJ gives authors a choice of peer review models based on a more or less conventional workflow: PeerJ "encourages" reviewers to identify themselves to authors, and authors will have the option of publishing their full peer review history alongside their published article.

My view is that only an open review process where it is mandatory for reviewers to disclose their identity, together with a review process that is open to all qualified scientists — at any time — properly fulfils the demands of transparency. This in turn leads to a higher degree of quality, which is vital for effective scholarly discourse.

Consequently, any limitations on a fully transparent, open PPPR process will hinder that goal and is essentially a compromise that — while it may make life easier for everybody in the short-term — cannot provide long-term openness and sustainability in scholarly communication. The point is that a maximum level of quality is essential for the high-level academic discourse the networked environment will require.

That is also why I do not believe it is enough to simply launch yet another new open access journal, or to flip existing toll-access journals to open access, or to start a mega-journal site (which is principally the same as starting a new journal but on a broader scale). It is likewise not sufficient to add a few new features to the peer review process that we have inherited. Doing so is likely to be more confusing than useful to the broader academic community.

The pure concept of public PPPR requires no such compromises and so is more likely to result in the radical transition of today's conventional publishing system to the system that will be needed in the future.

We have, by the way, also just introduced (during Peer Review Week, of which we were co-founders) a new peer review experiment called Peer Review by Endorsement (PRE) with author-led pre-publication peer review.

RP: You would not characterise ScienceOpen as a megajournal then?

AG: Since we publish in all categories, ScienceOpen Research could qualify as a megajournal, but at its current size the term "mega" might be somewhat of an overstatement! We are moving more towards offering publically available abstracting & indexing services, and powerful search, than a megajournal, which I can say more about.

RP: Ok, we can pursue that further. But first can you confirm that as well as the names of the reviewers, the reviews themselves are all publicly available on ScienceOpen? I couldn't find some of the reviews mentioned on [one of the pages](#) you directed me to. Maybe some are published as separate articles? I also assume that you are unlikely to attract reviews for papers that you have aggregated from elsewhere rather than published yourselves, or papers where you hold only bibliographic details. Would that be a fair assumption?

AG: I checked the links in the blog post and they seem to be working. The high profile of the reviewers has been highlighted and not in every case their actual reviews, which you can find on every article mentioned.

Here are the links: [Hugo Ten Cate](#), [Nitika Pant Pai](#), [Daniel Graziotin](#), [Nikos Karamanos](#).

As to attracting reviews for papers published elsewhere, that would be the desired state but it will require a big culture shift in the scientific community – one that we hope we can help facilitate by showing how helpful post publication reviews can be for the reader.

RP: How would you respond to someone who said: "If you believe we should be working towards PPPR perhaps there is no need for a service like ScienceOpen. It would be more cost effective simply to have researchers post their preprints to servers like [arXiv](#) and [bioRxiv](#) and then create overlay journals in the style of [Discrete Analysis](#), the journal being developed by [Timothy Gowers](#) This would save millions of dollars in Gold OA fees."

AG: Yes, I read about that idea and it's powerful. I think that there needs to be greater collaboration between all of us who are trying to reform the scholarly communication system and if overlaying ScienceOpen functionality directly on top of pre-print servers is what the community feels would be beneficial, then bring it on and let me know.

To be honest, it was that specific vision that first inspired me (as a physicist) in the Nineties to think about a new world of scholarly communication in which we could facilitate the speedy exchange of new concepts, ideas, experiments, (negative) results and their interpretation — and that's the kind of scholarly 'discourse' that we want to foster at ScienceOpen, as our Board Member at the Wellcome Trust Centre for Neuroimaging in London [Ray Dolan](#) described very accurately years later in a personal meeting with me.

Numbers

RP: Ok, let's look at some numbers. Can you share with me some statistics on ScienceOpen — e.g. the total number of papers it has published, the number of papers it hosts (because, as you say, you aggregate a lot of papers from third-party sites too), the number of reviews that have been posted, the number of hits on different areas of the site etc.?

AG: During the 12 months since we launched our publishing services we have generated 135 works (among them 55 research papers) which have been read by more than 85,000 users so far. Already about half of those research articles have received a peer review, some of them two or three, with 100 reviews published so far. We always expected a delay in the responsiveness of researchers to public PPPR during this introductory period.

We are particularly satisfied with the quality of the reviews that have been [submitted by reviewers](#). These compare well with our personal experiences (both as reviewers and editors) at journals that operate a closed peer review process.

As I said earlier, we have also added nearly 2 million (Gold) OA articles to our site, of which more than 800,000 are harvested from arXiv the largest preprint server worldwide, and [PubMed](#). Meanwhile we have indexed more than 100 million references and added those articles that have been cited to the site too.

Depending on the level of openness and completeness we describe our content in terms of article records. Only the minority of these records are open access but we can at least show that information is legally available to users who don't have subscription access to the relevant journal.

This is a great step forward in terms of improving users' experience when searching or browsing for content of relevance to them.

We can also use the references of those articles that have been made openly accessible after an embargo period to further build up our own, open indexing and citation database.

These activities are ongoing and we expect to achieve a data set of about 15 million articles by the end of the year. In parallel, we plan to set up an author profile database consisting of more than 5 million records for those authors we can identify via their existing profile at ScienceOpen or their record at ORCID.

Of course, our database is not yet complete but nevertheless we can support users with a high level of confidence in deciding if a publication is of higher relevance to them than a comparable work.

Meanwhile, our citation database summarises more information than is available elsewhere on a single site. For example, we provide the total number of times an article is mentioned on social networks. This service is provided to us for *all* 10 million articles by [Altmetric](#).

RP: Can I just check my understanding: ScienceOpen contains 10 million articles and references. Of these, 1.2 million are open access articles that you have harvested from OA publishers and from PubMed, and 800,000 are papers that have been pulled down from arXiv. The rest (8 million) are simply the bibliographic details of papers published in subscription journals. Is that right?

AG: That is basically correct however the numbers are slightly off: we have 950K open access articles imported from the [PubMed Central Open Access Subset](#) and 830,000 from arXiv. The

additional roughly 8 million bibliographic records are extracted from references within the full article content. We do not have licensing information about these bibliographic records, but I assume that most of them are from subscription journals.

We have started rounding out these records with the full metadata from external sources such as PubMed and ORCID to enhance the usability of the content. We also compare the bibliographic records to each other to ensure that we have good matching so that we can merge records.

Costs

RP: How much does it cost to publish an article with ScienceOpen?

AG: ScienceOpen charges an APC of 800 USD for an article regardless of its type or length. We do not publish monographs or books. Our new peer review experiment with author-led pre-publication peer review operates an APC of 400 USD (See here)

Posters can be published for free and receive a DOI. They show up in the same style in the database as any other record on ScienceOpen. All our pricing information can be found here. We also offer fee waivers to those who can demonstrate need.

RP: Who can get a fee waiver, and how many waivers has ScienceOpen given to date?

AG: We provide full and partial fee waivers to those able to show that they are unable to secure funding from their institution, an OA publication fund, or some other available source of funding.

Since launching we've offered fee waivers to early career researchers. We also allow Max Planck researchers to publish with us at no charge. Additionally, we have given Greek researchers the opportunity to publish at no charge until the end of 2015, to help ease their fiscal woes.

Finally, in September we awarded prizes to more than 20 early career scientists in the medical sciences from Europe and Asia on the occasion of an annual conference at Charité Medical University Berlin. These researchers were awarded a free publication of their work on ScienceOpen and inclusion in an ESC Poster Collection.

RP: You said you recently introduced a new Peer Review by Endorsement (PRE) experiment. This involves pre-publication peer review. Does this experiment not suggest that you believe pre-publication offers some advantages over PPPR? If so, what are these?

AG: It is always a good idea to have a colleague read through your paper and give feedback before making your results public. We hope that by encouraging researchers to take this step even better quality manuscripts will be produced.

Our experiment is just that, it's another option for researchers to try. In terms of advantages, I think it's a matter of personal preference. Some researchers may want scrutiny before publication and that's reasonable as long as it doesn't delay sharing the

results, others are happy just to publish and then revise their manuscript based on reviewer feedback (two revisions are included in our publication fee).

At ScienceOpen it's even possible to combine both Pre and Post Publication Peer Review processes, if a researcher is looking for absolute rigour.

RP: The PRE experiment invites scientists to choose their own peer reviewers. Is there not a danger that this will encourage them to try and cheat the system (by scratching each other's backs for instance). Are you confident that the rules you have put in place for your peer review by endorsement option will avoid cheating?

AG: I think so. All publishers (traditional and non) are interested in who a researcher would pick to review their work, since they know the landscape of expertise better than any professional editor ever could. We are simply transparently acknowledging this fact and encouraging authors to share their contacts in the interests of getting their publication reviewed faster.

We assume that the social and professional cost of giving good scores to bad research in an open review system would be too high to make cheating very attractive, which is one of the reasons that we insist on complete transparency. Cheating actually works better in the traditional anonymous system of peer review.

In terms of acknowledging this issue in the PRE methodology, we do have some clearly stated rules in place to help mitigate this and I do urge you to check them out, they are fair but firm!

Citation network

RP: Can you say more about your citation network?

AG: ScienceOpen intends to provide researchers of all academic disciplines with a range of contextual information on the articles of relevance to them. This will include records, names, citations, social mentions, and so on.

To achieve a working proof of concept, in the Spring we started to systematically track all the references of the nearly 2 million open access articles accessible on ScienceOpen as part of our original aggregation efforts (so PubMed Central, ArXiv and ScienceOpen itself). This is only the starting point for the further deployment of our platform.

My partner and co-founder of ScienceOpen, Tibor Tscheke, who has been familiar with these concepts and the publishing industry in general since the early Nineties, has summarised our vision for the next few years in a recent blog post.

If we compare the evolution of the media industry in general, we can see that the importance of network and service providers has evolved during the last decade. This has seen the established role of content providers and their brands begin to diminish.

If we then look at scholarly publishing, we see that every year about 2 million new articles are published in STM alone (out of 4 million submitted manuscripts). So even at its most successful, a new for- or non-profit venture can expect to have published only a few thousand to 10,000 papers after, let's say, five years. This is neither a significant market share, nor does it contribute to the disruptive change in the publishing industry that we can expect to see take place, and which has just started. As such, I see no reason to place undue stress on the publishing side of ScienceOpen, which can only expect to publish a few hundred articles in the first couple of years.

We are still committed to publishing open access research articles using an innovative model in our journal *ScienceOpen Research*, but our focus has shifted towards aggregation services. This is where we see that we can offer the scientific community new services that make research results more openly available and findable.

Researchers want to search the output of their peers no matter where and when that research is made available — be it posted on a preprint server or published in a journal or megasite. As more information becomes available, it will be possible to harvest that content and make searching progressively more useful as a result.

Centralise or federate?

RP: I read the article of yours you mentioned earlier. A number of questions occurred to me. For instance, I was not sure whether you believe the emphasis should be placed on centralised or federated aggregation services. You talk, for instance, about saving merely metadata on central services. As I understand, however, ScienceOpen's long-term goal is to store only full text. Is that right?

AG: As discussed, ScienceOpen has around 2 million open access full text articles and over 8 million metadata records. As open access publishing gains momentum, I expect that even more structured full text records will become available for public archiving. For the moment, almost 90% of research is still published behind a paywall. Opening up and adding context to these records is the first necessary step.

To help in this process we have started to invite publishers to provide us with their full-text content [for indexing purposes](). We have deployed the necessary features for this in the new release of the ScienceOpen platform, and so will be able to provide a much better search experience for researchers. This will also be of benefit to partner publishers by promoting their content.

With this new offering, we will be able to build on open access by providing *open indexing information* at the point of (re)search. It's great for both authors and publishers to see that their articles are presented in the context of other, similar articles from other publishers.

This is a completely new way of discovering scholarly content regardless of when or where it has been published. That's why we called the new functionality on ScienceOpen "Discovery".

When drafting the manuscript for a recent paper, I myself spotted some very useful articles on our platform, articles that I would never have otherwise found. This feature is unique as far as I know and will help foster demand for more open access at the content level.

RP: You mentioned that you harvest papers from arXiv. I am wondering if you are adding self-archived papers to ScienceOpen? If so, I suspect this might be problematic for publishers like Elsevier. Does this pose any problems for you?

AG: We are not adding self-archived papers to ScienceOpen. Rather, we are working with publishers to get the maximal amount of information surrounding an article onto ScienceOpen, including references, keywords, etc. From the references of each OA article we create new article records, and open these up to commenting and review.

So, for example, while we have a lot of article metadata from Elsevier and other big publishers on the site, unless a publisher works with us we will not link back to its website.

For open access articles we have more information and therefore our "related articles" function is more effective, giving them a discovery advantage over closed access article records.

RP: You say you don't link back to subscription publishers' websites. I noticed that the links to articles published in journals like Nature *and* Cell *not only do not link back, but the records in your database appear to consist of no more than the authors' names and the publication date, often with no title. Sometimes the authors' names are missing too. What value is ScienceOpen providing to users with these kind of records, which I would imagine are the majority of the 10 million records you offer?*

AG: Because we ingest data from various sources the initial quality of the reference extraction will be dependent on the consistency of the data. The data produced by different publishers varies in completeness and consistency (the reference style of some journals for example leaves out title or gives only the first 3 authors) and we are working to fill the gaps with other services to increase the quality of the content on our site.

We are also adjusting our algorithms to reprocess inconsistent/incomplete data to provide a better experience for our users on an ongoing basis.

We are at an early stage and already can offer users very good discovery tools with our similar article recommendations based on title words and author similarities even from these records.

And we can offer a ranking of articles according to relative citation count. The ability to filter a search based on citation count is currently only available in tools such as Web of Science and Scopus through expensive library subscriptions.

Moving forward, we would like to work actively with publishers to improve the quality of their content on ScienceOpen, add more references and link back to the publisher page to

generate traffic at the original article source. Through this we can provide a valuable service for both researchers and publishers.

RP: In your article you argue that there is no longer any need for traditional journals and this seems to be the thrust of your argument in this interview. The founder of F1000 Vitek Tracz agrees. _Speaking to me recently_, he said: "[The] main problem [with scholarly publishing today] is the very existence of journals, and the methods their editors use to accept or reject what to publish. Their reason for making these choices is significantly affected by their battle to maintain and raise their Impact Factor." He added, "no one reads journals, everyone reads articles."

But traditional publishers appear to see things very differently. For instance, Robert Harington, Associate Executive Director, Publishing, at the American Mathematical Society has _argued_: "[T]he journal is a matter of record, and like it or not, journals form a part of the academic and recognition workflow that allows for career progress, grant making, more research and more articles to be published … What a journal represents to the community is more than just the pages themselves. How else over the span of many years can you keep up with the corpus of human endeavour?" How would you respond to Harrington? Is it really likely that the research community will soon agree to give up the journal as we know it?

AG: The way researchers themselves are evaluated maintains the over-reliance on journals. The insidious way the Impact Factor of the journal in which a researcher has published still influences their career prospects is untenable.

Until the research community decouples the journal from the article (so that we value the latter more), this problematic cycle will continue. However, the revolution has started, and you can see the evidence in the DORA [San Francisco Declaration on Research Assessment] initiative. Moreover, it seems to me that each year the noise generated by Impact Factors diminishes (or perhaps this is just wishful thinking on my part!).

What I would really like to see is for research funders to push forward in this space, because money follows money whether we like that or not.

And I do completely agree with Vitek's statement: "no one reads journals, everyone reads articles." This is why we created the concept of the Collection at ScienceOpen. We feel that it is becoming less and less important where (and also when) an important finding is published.

What is necessary is that it is easily findable and accessible, and a Collection is a great way to aggregate those findings that have been published elsewhere, in a curated list of articles.

The company

RP: Can you say something about the company itself? Where is ScienceOpen located (i.e. in which jurisdiction has it been registered and where will it file its financial accounts) and who are the shareholders?

AG: ScienceOpen launched in July 2013 in Berlin as a GmbH (comparable with a "Ltd" in the UK) and it is still privately owned by the two founders Tibor Tscheke, based in Boston, and me (based in Germany).

In parallel to that launch, and to reflect the distribution of intercompany services, we launched ScienceOpen, Inc., in Boston, US in 2013, too. This is 100% owned by ScienceOpen GmbH.

RP: You said ScienceOpen has published just 55 papers. I would think that some of these will have attracted an APC but perhaps most will have had the APC waived? I assume, therefore, that you are not earning much revenue today. As your focus shifts from publishing to aggregation, I am wondering what your business model will be going forward. How does ScienceOpen expect to earn revenue, and what are your revenue projections for the next few years?

AG: You would assume correctly. As you surmise, ScienceOpen started its product journey with three core services: publishing, aggregation and engagement, all of which are rooted in the principles of open access and transparency.

We discovered that users found our aggregation services to be the most compelling part of our offering and we have attracted attention for our focus on user engagement (Post-Publication Peer Review and Collections of articles from multiple publishers.) As I noted, we plan to focus on these two areas going forwards.

As discussed, we have [recently launched](#) our publically accessible citation data, which is free for researchers to use wherever they are (not just from one library terminal as now at our university) and is provided at no cost to libraries, liberating their funds for more creative endeavours. This initiative will be funded by modest indexing fees and marketing opportunities for publishers. It legitimises content sharing and increases traffic to their content.

Open access content will automatically be prioritised in search results and amplified in our database simply because those publishers have already made the full text available and so the links between it and other recommended content are stronger.

RP: I believe you also have a position as Professor of Publishing Management at the [University for Applied Sciences in Leipzig](#). What does that job involve, and how do you distribute your time: 50% at the University and 50% with ScienceOpen, or what?

AG: This is a very good question and my answer reflects my personal commitment to both positions. On the one hand there is no 50:50 split between these two roles because I have to give more than 100% of my effort to both jobs!

On the other hand, ScienceOpen is operated independently of me by a professional team. The team is responsible for dealing with all aspects of the business, including management, marketing, technical development and deployment etc.

Of course, it is obvious that as one of the owners, I am involved in all strategic issues and questions.

When I worked in the publishing industry as a managing and publishing director I learned how to maintain a high level of commitment to doing the (hopefully) right things and to making the right decisions. To come to the right decision in a very limited time frame is essential — and hopefully one makes the right ones!

I benefit a lot from my previous experience and I always question myself carefully when making decisions. I can also check in with my colleagues and others in order to get their opinion but, fundamentally, if a decision feels like it can't be readily explained in a transparent manner, such as on this blog, then it's probably not the right decisions!

All of this is physically and intellectually extremely demanding for an individual and it only works if you receive positive feedback and have fun in what you are doing despite the huge time, effort and money you are spending in order to reach the next step of your goal.

It reminds me of soccer or, as we call it in Europe, football: you can only win if you continue to believe that you can do. And if you have some fun while doing it as well that is just great. I would add that with start-ups only those in the game for the long haul can hope to win.

I play soccer myself once or twice a week and as long I can continue with this I am confident that — with the outstanding team of enthusiastic and experienced people at ScienceOpen (both here in Germany and in the US) — we can achieve our common goals.

RP: Do you think there are any potential conflicts of interests in being a professor of publication management while simultaneously running a commercial publishing operation? If so, how do you manage these conflicts?

AG: For the most part, this issue generally doesn't arise, so long as I continue to separate both activities in a professional way.

I can differentiate between my views as the individual who created the venture ScienceOpen and those of somebody who works in academia with a clear focus on teaching students in publishing management.

The state of open access

RP: How would you say the environment for open access has changed over the past year? What positive developments would you point to, and what setbacks would you draw attention to? What is the state of open access today?

AG: I summarised the current aspects and relevance of open access as I see them in the interview I did two with you years ago. If we look at the broader picture, in my opinion nothing much has changed since then. Certainly open access has moved forward very little.

It is doubtless for this reason that business analysts concluded several months ago that the "threat" of open access for legacy publishers has disappeared. However, I am quite confident that they are completely wrong.

When I talk to colleagues in the publishing industry it is perfectly clear to me that they are aware something is going to have to change drastically in the near future. The process of a so-called 'disruptive' transition has only just started and we will bear witness to the process of change in the coming years.

As a result of the developments I described above I predict that — with the exception of maybe a few highly regarded journals — most scholarly titles will soon begin to lose their influence. And more and more scholars will demand open and transparent peer review in order to improve the quality of the assessment process that academic research has to undergo.

Above all, the necessary changes all demand immediate open access. The majority of researchers and the major publishing houses now realise that.

Yet you will recall that only a few years ago publishers were deeply opposed to open access and lobbied against it whenever they could. In doing so they peddled tales about the poor quality of open access journals — a general statement that is complete nonsense. When I started working in the publishing industry I heard these tales repeatedly, and sadly some researchers still have them in mind when they express views on open access today.

The situation started to change in 2008, when Springer decided to acquire Biomed Central. I was working at Springer that year and completely understood the rationale behind the acquisition as it was summarised by CEO Derk Haank in a management meeting. Although the wisdom of the acquisition was not self-evident to everyone at the time — not least because it was far from certain that the future would be predominantly open access — today we can see it to have been a smart investment.

If we look at the number of open access articles that are being published by the key players in academic publishing today, we observe that Springer has become the pacesetter and market leader in (gold) open access output, especially after its merger with Nature. Other key players are Hindawi and PLOS.

Lost ground

By contrast, major publishers like [Wiley](#), [Elsevier](#), and [Taylor & Francis](#) appear to have lost ground and will have to improve strongly if they want to keep on track.

Unfortunately, the weakest link in the open access movement today appears to be on the other side of the market. I'm talking about universities, libraries, and those other organisations and individuals who should by rights be highly committed to open access but are not.

When attending open access conferences, or speaking to these individuals, I get the feeling that their discussions have stagnated and become more focused on the technical aspects or details of OA rather than on devising and establishing new ways of boosting the necessary transition to OA.

This is not just my observation. Recently, a plenary speaker from Brussels asked the audience at the OA Days conference in Zurich if they wanted to continue to talk to each other or do something positive for their users.

It is an odd situation because those who have access to the huge budgets for the acquisition of academic literature seem to be waiting for something to turn up (perhaps an alternative to APCs) rather than actively supporting the deployment of new ideas and methods for fostering gold open access, and speeding up the transition.

When I ask librarians if they are supporting OA at their institution or library, they tend to respond that they will need additional money in order to do so.

I find this answer somewhat frustrating because every year more than 10 billion USD is spent on subscriptions to scientific, technical and medical (STM) journals alone. Obviously, therefore, there is quite enough money in the system and it would be easy to deploy this in order to ensure (much) more OA is provided. Those who have been responsible for these huge budgets for decades know this much better than any of us.

Anyway, the good news is that there is no shortage of great ideas out there, and many ways in which we could proceed quickly if we chose to. We really don't need to wait another 5 or 10 years for the transition. Just let's go for it, now!

Copyright

RP: You are clearly in a hurry for open access to become the norm in scholarly communication. In this regard, I was struck by what Barbara Fister *said in a recent* Inside Higher Ed *column. Pointing out that historically authors always assigned (and many continue to assign) copyright to publishers, she concluded that the bulk of the research corpus will not be released from paywalls in either her lifetime or her children's lifetime. This surely significantly limits what you will be able to achieve with ScienceOpen does it not?*

AG: The thorny issue of traditional copyright and its role in stifling scientific discourse between and beyond peers does indeed need addressing. I've recently enjoyed reading a number of articles and blog posts about how authors can push back on publishers who want them to hand over their rights and I think explaining this clearly to them is important.

Of course, it's not up to authors to fight this battle alone and each and every funder mandate is essential in this regard, since it is by this route that we can finally liberate research. As I say, rightly or wrongly, money follows money. I am also confident that publishers will eventually consider releasing the copyright they hold in scholarly content so that it can be made more openly accessible.

You know from [your last interview with me](#) in 2013 that I am not a fan of 'Green' OA but for that huge bulk of research content published in former decades it could at least be the first step in a period of transition from toll to open access.

RP: I have to say I am somewhat sceptical that publishers will voluntarily give up their ownership in research papers. Doubtless I will be damned as an incorrigible socialist (or even a communist) for asking this, but do you think perhaps governments should step in to free this information? Might there be an argument for saying that all research papers currently held behind paywalls should be released into public ownership, in the way some industries were nationalised [in the UK after the last war](#)?

AG: I think the phrase Bernie Saunders uses is "democratic socialism" and I am right there with you in pledging allegiance to that cause.

Seriously, when you stop and consider that some seminal articles that were written many moons ago still cost $35 or more to download as a PDF, then really I could weep for the nonsense of it all!

The impediment to scientific progress caused by our current regressive publishing system, and the slow pace of change, are two of the few things that keep me awake at night. That said, usually the combination of my family and soccer wipes me out!